

Improved analysis of anomalous diffusion data in single particle tracking experiments

Eldad Kepten¹, Yuval Garini¹,

1 Physics Department & Institute of Nanotechnology, Bar Ilan University, Ramat Gan, Israel

*** E-mail: Corresponding Eldad.Kepten@biu.ac.il**

Abstract

The Mean Square Displacement is a central tool in the analysis of Single Particle Tracking experiments, shedding light on various biophysical phenomena. However, as we show, it suffers from two systematic errors when analysing tracks of anomalous diffusing particles. The first is significant at short time differences and is induced by measurement errors. The second arises from the natural heterogeneity in biophysical systems. We show how to estimate and correct these two errors and improve the estimation of the anomalous parameters for the whole particle distribution. As a consequence we manage to characterise ensembles of heterogeneous particles even at very short and noisy measurements where regular time averaged mean square displacement analysis fails. This procedure has the potential to improve experimental accuracy while maintaining lower experimental costs and complexity.

Notation

Anomalous diffusion constant - D_α
 Apparent anomalous exponent due to heterogeneity - α_S
 Apparent anomalous exponent due to noise - α_N
 Apparent location - $\hat{x}_{(t)}$
 Apparent TAMSD - $\hat{\delta}_\Delta^2$
 Control experiment noise MSD - \tilde{N}_c
 Dynamic functional - φ
 Error in anomalous exponent due to heterogeneity - d_α
 Mean Logarithmic Square Displacement (MLSD) - $\lambda_{(\Delta)}$
 Mean of anomalous exponent distribution - μ_α
 Noise and heterogeneity corrected MSD - $\nu_{C(\Delta)}$
 Noise Corrected MSD (NC-MSD) - $\nu_{(\Delta)}$
 Noise MSD - \tilde{N}
 Relative error caused by noise - ϵ_N
 Standard deviation of anomalous exponent distribution - σ_α
 Standard deviation of noise - ρ
 Time averaged MSD - $\overline{\delta}_\Delta^2$
 Time difference - Δ
 True location - $x_{(t)}$
 True particle anomalous exponent - α

Introduction

Single Particle Tracking (SPT) is an increasingly popular technique in the study of biophysical systems [1]. Typically, a single particle is tracked as it moves and its trajectory is extracted. This measurement can

be performed both in vivo or in vitro. Advances in biochemical labelling, nano particle production and the use of fluorescent proteins now allows to label almost any biological entity such as, single stranded RNA in bacteria [2], nano-beads in a polymer solution [3], membrane dynamics [4], and specific nuclear entities [5, 6].

At each time point, the location of the particle is measured and stored for later analysis. In most biophysical systems, the motion of the tracked particle is stochastic - thus further analysis is necessary before claims regarding the underlying nature of the system can be made.

Perhaps the most basic analysis technique is the Mean Square Displacement (MSD) in which one calculates the average square of the distance $x(t)$ past by the particle in time t . This can be performed in two ways. The first is the ensemble averaged MSD (EAMSD), in which the averaging is performed over the displacements of N particles, at some t since the beginning of the experiment, $\langle x^2 \rangle_{(t)} = \frac{\sum x^2(t)}{N}$. The second is the time averaged MSD (TAMSD) in which all displacements over a time difference of Δ from a single trajectory are averaged over:

$$\overline{\hat{\delta}_{\Delta}^2} = \frac{\sum_{i=0}^{T-\Delta} [(x_{(i+\Delta)} - x_{(i)})]^2}{T - \Delta} \quad (1)$$

Stochastic processes can be characterized according to their MSD behavior, be it the EAMSD or the TAMSD. If the MSD is linear in time, i.e $\langle x^2 \rangle = Dt$, then the process is termed normal diffusion, where D is the diffusion constant. This is the behavior of, for example, Brownian motion. Other MSD behaviors are termed anomalous diffusion [7, 8].

If the MSD is described by a power law, $\langle x^2 \rangle = D_{\alpha} t^{\alpha}$ then a further distinction is made. When $1 > \alpha > 0$ the process is called Subdiffusion and for $\alpha > 1$ the process is called super diffusion. D_{α} is the anomalous diffusion constant and α is the anomalous exponent. We stress that the TAMSD and the EAMSD do not have to show the same behavior, and that the behavior of a process may change depending on the analysis time t (or time difference Δ) [9].

The physical and mathematical origins of anomalous diffusion are diverse. The most popular models are the Generalized Langevin equation (GLE) and Fractional Brownian motion (FBM) which can describe the motion of many body systems and viscoelastic media [10–14]; Continuous Time Random Walks (CTRW) which can describe biological trapping [9, 15, 16]; and fractal landscapes which describe, for example, highly obstructed motion [17, 18].

The MSD itself cannot be used to distinguish between these models alone as they can all lead to similar anomalous exponents. This raises a significant challenge when analysing biophysical data. Since in many systems SPT is applied in order to probe the dynamics and physical mechanisms governing the system, the ambiguity of the MSD results hinders clear conclusions. However, with the advancement of mathematical research, more subtle tests have been proposed which succeed in identifying the nature of the stochastic process behind the random motion [19–23]. Some of these tests have been implemented successfully to experimental and simulated systems [4, 13, 24, 25].

Another important step when studying the biophysical system causing the stochastic motion is the extraction of accurate parameters that characterize the system. Accurate parameters can shed light on the physical process taking place. For example, the dynamics of various polymer models can be described by a GLE model, but each will give rise to a different anomalous exponent. As an example, a phantom Rouse chain will have $\alpha = 0.5$ while a Zimm model gives rise to $\alpha = 2/3$. Various chain properties can alter this value [26], and in the long time domain, all chain models behave similarly according to the crowding of the media [27]. Thus, by identifying both the mathematical model and the correct anomalous parameter, one can gain deeper understanding of the physical process.

This task is not a simple one. Biological systems are characterized by heterogeneity of both the tracked particles and the surrounding medium. In addition, the total time that the biological system can

be measured is limited due to photo-bleaching, induced photo-damage to the system and cell migration. This inserts a variation of the motion parameters between the various particles, even in the same system.

Since the system is heterogeneous, the average tracer dynamics and the spread of parameters is important. If the single particle TAMSD is noisy, and cannot be analysed accurately than the ensemble averaged TAMSD, $\langle \delta_\Delta^2 \rangle$, must be calculated and fitted to its parameters, hopefully reproducing the average behaviour. In this averaging, many more data points are available for each time difference than by taking the time average of each particle alone. Evidently, random errors are significantly reduced. However, one cannot accurately assess the distribution of parameters with this technique. The assessment of this distribution is essentially left dependant on the accuracy of individual TAMSDs.

This challenge in the assessment of individual parameters has been found regarding the diffusion constant in normal diffusion [28–30]. In these studies, the distribution of fitted diffusion constants and their dependence on measurement rate, track length and more have been developed. However, to our knowledge, no such framework has been developed for anomalous diffusion. Specifically, an understanding of the systematic errors of the above techniques is missing.

In what follows we take an experimentally applicative approach regarding the accurate extraction of the distribution of anomalous exponents from experimental data. We identify two systematic errors that are expected to arise in many biophysical SPT measurements. The first is an offset due to the limited spatial precision (location accuracy) and the second is due to the heterogeneity of the individual particle's behaviour.

As we show, both can be estimated much better by using common mathematical methods. We introduce two improved measures that substitute the standard MSDs: 1) the Noise Corrected MSD (NCMSD) and 2) the Mean Logarithmic Square Displacement (MLSD). With the combination of these two measures, the average and standard distribution of anomalous exponents can be estimated correctly. In addition, the average particle TAMSD can be calculated to give the average diffusion constant. Simulations of anomalous motion illustrate both the systematic errors that occur while using simple MSD averaging and the implementation of the improved method.

With these simulations we show how to characterise very short and noisy trajectories, an impossible task with regular MSD analysis. As we explain in the conclusions, this new capability has the potential of reducing experimental costs and complexity while at the same time improving the accuracy of the results.

Discussion and Results

Short time α deviation

Every measurement has inherent noise and uncertainty. In SPT measurements, the location $x_{(t)}$ can only be measured with a varying noise term $N_{(t)}$ that depends on various factors including the point spread function (PSF) of the optical system, the acquisition rate, and other optical and mechanical limitations [30, 31]. In general, $N_{(t)}$ is a random variable and its distribution dependent on t and $x_{(t)}$. However, we consider the simpler case for which the noise is not dependent on time, i.e $P_N(t) = P_N$. This is the common situation for good SPT experimental conditions - negligible tracer bleach and constant spatial optical limitations.

We further assume, based on the central limit theorem, that $P(N)$ is a normal distribution with zero mean and ρ standard deviation. Thus each measurement is actually the sum of the true position and the normally distributed noise: $\hat{x}_{(t)} = x_{(t)} + N_{(t)}$. The measured TAMSD is therefore:

$$\overline{\hat{\delta}_\Delta^2} = \frac{\sum_{i=1}^{n-\Delta/\tau} [(x_{(i+\Delta/\tau)} - x_{(i)}) + (N_{(i+\Delta)} - N_{(i)})]^2}{[n - \Delta/\tau]} \quad (2)$$

Where n is number of time points along the trajectory, Δ is the size of the sliding window for averaging and τ is the acquisition rate.

The true process and the noise are uncorrelated. In the case where Δ is small and there are many averaged points, the cross terms between the two differences is eliminated. Also, the noise is independent of time and its MSD is therefore equal to $\tilde{N} = \rho^2$. Thus we are left with $\overline{\delta_\Delta^2} = \delta_\Delta^2 + \tilde{N}$, where δ_Δ^2 is the MSD of the true process. For an anomalous process with diffusion constant D_α and anomalous exponent α this gives $\overline{\delta_\Delta^2} = D_\alpha \Delta^\alpha + \tilde{N}$.

In order to extract the anomalous exponent one calculates what is commonly termed the dynamic functional, φ , by taking the derivative of the logarithm of the MSD according to the logarithm of time, $\varphi = \frac{\partial \log \overline{\delta_\Delta^2}}{\partial \log \Delta}$. For sub- and super-diffusion, one expects $\varphi = \alpha$. However, in the case described here, the added noise term will give an erroneous anomalous exponent, $\varphi = \alpha_N$:

$$\alpha_N(\Delta) = \alpha \frac{1}{1 + \tilde{N}/(D\Delta^\alpha)} \quad (3)$$

The relative error of the measured dynamic functional is $\epsilon_N = (\alpha_N - \alpha)/\alpha$. After a short manipulation we find:

$$\epsilon_N = - \frac{1}{1 + (\overline{\delta_\Delta^2} - \tilde{N})/\tilde{N}} \quad (4)$$

From equation 4 and 3 several facts emerge. First of all, the noise always works to lower the anomalous exponent. This is in line with previous findings which have shown a lowering of the anomalous exponent in simulations [32]. Second, the error is time dependent. At very short time intervals, $\Delta \rightarrow 0$, we find $\epsilon_N \rightarrow -1$ and $\alpha_{N(\Delta)} \rightarrow 0$. At long averaging intervals, the error becomes negligible and $\alpha_N \approx \alpha$.

Figure 1A shows the TAMSD of a single particle with $\alpha = 0.5$ and with noise of magnitude $\rho = 1.2$. The deviation from the theoretical prediction is seen clearly at small Δ values. Notice that as this is a single particle track, it is not smooth. The prediction of equation 3 is presented in figure 1B together with the average dynamic functional calculated from ten individual tracks and a clear fit is seen. We stress that averaging a large quantity of tracks will not eliminate this systematic error as it does not arise due to limited sampling.

Luckily, this systematic error can be approximately corrected with simple experimental measures. Suppose a control measurement is performed in order to assess the magnitude of the noise. This is actually a necessary step in any experimental system and can be performed by various means. In live cell measurements, for example, the measurement noise is sometimes classified by measuring a chemically fixated cell [5]. The static noise data can then be analysed to give the MSD of the noise alone - \tilde{N}_c which should be constant in time. Note that this measurement does not estimate effects such as motion blur. However, these are typically much smaller in SPT experiments, especially with short exposure times [30].

After finding the MSD of the noise, one can calculate the Noise Corrected MSD (NC-MSD), $\nu_{(\Delta)} = \overline{\delta_\Delta^2} - \tilde{N}_c$. Since the control experiment is assumed to give the same noise as in the actual measurement, one finds $\nu_{(\Delta)} \approx D\Delta^\alpha$. Now, calculation of the dynamic functional will give α with only negligible error.

This correction is applied to the previous simulations and presented in figure 1A and B. It is clear that the NC-MSD accurately retrieves the anomalous exponent of the simulated process even when the noise is of the same magnitude as the process MSD. This experimental scenario was previously impossible to study without large errors. From the theoretical point of view, the NC-MSD enables the measurement and analysis of anomalous diffusion down to unlimited short measure rates.

Large time α variation

When calculating the TAMSD of a single particle, for large Δ values or short tracks, there are not many points to average over. Hence the TAMSD suffers from high variation and random errors [28]. These errors are obvious for example in Figure 2A where it is impossible to analyse the TAMSD after the first few time points.

A possible way to mitigate this problem in the case of normal diffusion is to take the ensemble average of many TAMSDs (EA-TAMSD) of different tracks, $\langle \overline{\delta_\Delta^2} \rangle$. This method works adequately for normal diffusion or for identical particles, retrieving accurate average diffusion parameters for the system.

However, for the general case of anomalous diffusion, where the tracers and their environment are not identical this method fails. In other words, the ensemble average TAMSD does not represent the behaviour of the typical particle. As we show shortly, the dynamic functional extracted from $\langle \overline{\delta_\Delta^2} \rangle$ is different from the average anomalous exponent and shows a time dependency.

Let us begin with the simple case of two particles. We give both of them a diffusion constant of $D_\alpha^{1,2} = 1$ and assume $\alpha_1 > \alpha_2$. It is clear that the average anomalous exponent is $\frac{\alpha_1 + \alpha_2}{2}$. However, the average TAMSD is $\frac{\Delta^{\alpha_1} + \Delta^{\alpha_2}}{2}$. Upon extracting the dynamic functional from this average TAMSD, one will find that at long times, i.e. $\Delta \gg 1$ the average MSD behaves according to $\varphi \approx \alpha_1$ while at $\Delta \ll 1$ the behaviour is governed by $\varphi \approx \alpha_2$.

We now analyse a large ensemble of particles and take the common experimental situation where for each i 'th particle there is a different underlying anomalous exponent, $\alpha_i \sim P(\alpha)$. For clarity and simplicity, we take P to be a distribution with mean μ_α and standard deviation σ_α . We note that since α_i must be positive, such a distribution must be limited at zero, i.e. $\alpha_i > 0$. However, this limitation is negligible unless $\sigma_\alpha \simeq \mu_\alpha$. The relevant corrections of the following equations for this scenario are presented in the SI. We also take $D_i \sim P(D)$ with an average \hat{D} . However, $D_{\alpha,i}$ and α_i are not correlated. As we will see later, the diffusion constants do not matter in the case of accurately extracting the anomalous exponent. This stems from our interest in the power law behaviour in long and short times, dimming the diffusion constants influence negligible.

For the above distribution of anomalous exponents, with each trajectory behaving according to the theoretical prediction $\overline{\delta_\Delta^2} = 2D_{\alpha,i}\Delta^{\alpha_i}$, then the ensemble average TAMSD behaves according to:

$$\langle \overline{\delta_\Delta^2} \rangle = \hat{D}\Delta^{\mu_\alpha} \cdot \exp[(\sigma_\alpha^2 \ln(\Delta)^2)/2] \quad (5)$$

It is clear that there is a time dependent offset from the average anomalous exponent μ_α . In order to find the time dependent anomalous exponent, $\alpha_S(\Delta)$, we again calculate the dynamic functional ($\varphi = \alpha_S$). One finds an effective average anomalous exponent of:

$$\alpha_S = \mu_\alpha + \sigma_\alpha^2 \log(\Delta) \quad (6)$$

From equation 6 we find the difference in the dynamic functional between two time lags. For Δ_1 and Δ_0 , and defining $d_\alpha = \alpha_{(\Delta_1)} - \alpha_{(\Delta_0)}$ we receive:

$$d_\alpha = \sigma_\alpha^2 \log\left(\frac{\Delta_1}{\Delta_0}\right) \quad (7)$$

Notice that d_α is independent of the diffusion constant and shows a logarithmic dependency on Δ_1 .

To test this result, 1000 particles with a distribution of individual anomalous exponents according to $\mu_\alpha = 0.5$ and $\sigma_\alpha^2 = 0.2$ were simulated for 1000 time points. The retrieved distribution of anomalous exponents is shown in Figure 2B. Two graphs of the MSDs of individual particles are shown in Figure 2A. The EA-TAMSD of the ensemble is shown in Figure 2C. Notice that the time difference and MSD are plotted on a logarithmic scales and a sub-diffusive MSD is presented as a straight line with slope α . Evidently, the EA-TAMSD has a changing slope, as expected from equation 6. The dynamic functional

extracted at each time point is seen in Figure 2D. It is clear that the error in the dynamic functional grows linearly with the logarithm of the time differences. Notice that these simulations were performed without the addition of measurement noise, and thus the single source of systematic error is the spread of anomalous exponents.

The time dependency of the dynamic functional is the attempt to compare between different powers of time. Hence, one should try to directly average the exponent values while cancelling the exponential time dependency of each trajectory. This can be done by averaging between the logarithm of the exponential quantities - creating a linear dependency in time. For this purpose we present the Mean Logarithmic Square Displacement (MLSD):

$$\xi_{(\Delta)} = \log\left[\sum_{i=1}^{T-\Delta} \frac{(x_{i+\Delta} - x_i)^2}{T - \Delta}\right] \quad (8)$$

For a process that shows anomalous TAMSDs, The MLSD gives for each particle i , $\xi_{(\Delta)} = \log[D_i] + \alpha_i \log[\Delta]$. Now, when taking the ensemble average, the true average exponent is found -

$$\langle \xi_{(\Delta)} \rangle = \langle \log(D) \rangle + \mu_\alpha \log(\Delta) \quad (9)$$

And specifically, $\varphi = \mu_\alpha$ for all times. The results of the MLSD for The same ensemble of particles in Figure 2B is also presented in Figure 2C and D. The perfect reconstruction of α_m can be seen.

Extracting the standard deviation of α_i

The MLSD gives more details about the anomalous behaviour of the diffusing particles than just μ_α . One can also extract the standard deviation of the individual anomalous exponents, σ_α .

At each time point, Δ_i where the slope of the MLSD is taken $\alpha(\Delta)$, one also takes the slope of the regular MSD, $\alpha_S(\Delta)$. Then, extract a set of the errors $\epsilon_i = \alpha(\Delta_i) - \alpha_S(\Delta_i)$ in estimation of the average anomalous exponent at times Δ_i . This set can then be fitted against $\log(\Delta_i)$ according to a small manipulation of equation 7, $\epsilon_i = \sigma_\alpha^2 \log(\Delta_i) - \sigma_\alpha^2 \log(\Delta_0)$. The slope of this fit, is the variation of the distribution of anomalous exponents in the population of particles, σ_α^2 .

It is important to notice two central points regarding this technique. The first is that both the MLSD and MSD are effected similarly by the noise induced error, ϵ_N . Thus the Δ_i 's used for the fit can extend to small Δ values, even though ϵ_N might be significant. This greatly increases the amount of points that can be used for the fit.

Second, in some experimental scenarios, this technique can be much more accurate than fitting each individual TAMSD to a power law. For example, if only a small amount of time points are measured, the individual TAMSD is very noisy. Thus the ability to extract individual anomalous exponents from each track is very low.

In order to test this technique, 5000 particles with α_i taken from a normal distribution with mean $\mu_\alpha = 0.6$ and standard deviation $\sigma_\alpha = 0.15$ were simulated. To further complicate the analysis, a measurement noise of $\rho = 1.5$ was incorporated into the process. For each track, only 2^6 time points were taken, thus the ability to fit each anomalous exponent is very low. Figure 3A shows the histogram of the simulated α_i values and the fitted ones. Two errors are evident. The first is the very wide spread of the fitted distribution, and the impossible negative values found in some cases. Actually, the standard deviation of the fitted α_i 's was $\sigma_\alpha^{fit} = 0.36$, an error of over 100%. The second error is the shifting of the peak value due to the measurement noise to an average value of 0.4.

Figure 3B presents the extraction of σ_α with the use of the MSD and MLSD. The measurement noise causes the dynamic functional found for the MLSD to be well below the average value of $\mu_\alpha = 0.6$. This however does not effect the result, as the standard deviation of alpha values was found to be $\sigma_\alpha^{MLSD} = 0.16$, an error of only 7%. We see that the MLSD-MSD difference technique extracts σ_α even for short, noisy trajectories, a task not possible previously.

Combining both techniques

With the tools of the NC-MSD and the MLSD, one may construct the average particle MSD from the measured trajectories. In doing so, one must take into account that the NC-MSD is affected by the distribution of anomalous exponents and will suffer an error according to equation 5, dimming it inadequate for large Δ . Since the MLSD cannot be easily corrected for measurement noise (as the logarithm cannot be negative) it will suffer an error according to equation 4 and give $\langle \text{Log}(D_i) \rangle$ instead of $\langle D_i \rangle$.

The key is to extract σ_α from the distribution induced error with the use of the MLSD-MSD differences as described above. Then, since we obtain an estimation for equation 7, the NC-MSD can be corrected to the true average dynamic functional by division.

$$\nu_c(\Delta) = \frac{\nu(\Delta)}{d_\alpha} \quad (10)$$

This procedure was implemented for a set of 1000 particles, measured at 1024 time points, with $\mu_\alpha = 0.6$ and $\sigma_\alpha = 0.2$. In addition, a measurement noise of $\rho = 1.5$ was added. Here too, the analysis of individual trajectories is bound for random errors.

First, the ensemble distribution error, d_α was extracted by comparing the MSD to the MLSD as stated above. Then, the NC-MSD, $\nu(\Delta)$, was calculated for each time point. Finally, equation 10 was used to find $\nu_c(\Delta)$. The uncorrected EA-TAMSD, $\langle \overline{\delta_\Delta^2} \rangle$ and $\nu_c(\Delta)$ are both presented in Figure 4A with the theoretical average particle MSD. The significant improvement is clear. Finally, Figure 4B shows the dynamic functional at each time point of both the uncorrected MSD and the final noise and variation corrected MSD. Evidently, equation 10 reproduces the average particle MSD accurately - both the average diffusion coefficient and average anomalous exponent are reproduced.

Conclusions

In this work we have studied two types of systematic errors in the estimation of anomalous diffusion MSDs, that may hinder attempts to accurately characterise such diffusive systems. These errors cannot be corrected by increasing the amount of averaged particle locations (either by longer measurements or more particles). Rather, they are inherent to noisy, heterogeneous systems. Such systems are the common situation in biophysics. As discussed in the introduction, such errors can lead to erroneous conclusions about the physical and biological nature of the system.

Every experimental measurement has a finite accuracy and incorporates some degree of uncertainty. In SPT experiments, the location of the tracked particle is known up to a random noise term. As we have shown, this added noise causes a systematic decrease of the measured anomalous exponent, even when the true MSD is much larger than the noise. For example an MSD to noise ratio of 4 will give an error of 20% in estimation of α (equation 4). We have shown how to quantify and correct this error with simple control experiments.

Heterogeneity in biology may stem from a variety of origins. At the top most level, there is variation between individual cells. Even inside each cell, each tracked particle may experience different surroundings, especially in finite time measurements. In addition, biological tracers may have differences in their properties, be they artificially inserted, artificially expressed or naturally occurring tracers.

Many times, the objective of the experimentalist is to characterize the distribution of parameters in a system. This is also true regarding the anomalous exponents of the individual particles. However, when analysing short tracks, it is difficult to accurately identify the parameters of single particles individually, since the small amount of time points introduces large errors. Hence, one aspires to reach conclusions from studying ensemble averaged quantities such as the EA-TAMSD.

We have shown that the variation of anomalous exponents creates a time dependent systematic error in the extraction of the average anomalous exponent from the EA-TAMSD. By using the MLSMD, we are able to correct this error and extract the true mean of the distribution. In addition, we study the propagation of the errors in the MSD, and manage to quantify the variance of the whole distribution of anomalous exponents.

Finally, we have shown how to combine the noise and the ensemble heterogeneity corrections. Thus an accurate MSD of the average particle behaviour is achieved (i.e. the average diffusion coefficient and anomalous exponent). This corrected average MSD suffers less random errors due to the large amount of points averaged upon and does not suffer the systematic errors of the original EA-TAMSD. Since this technique gives both the average anomalous exponent and its variance, the whole distribution of anomalous exponents is characterized.

This ability to characterise the whole distribution from ensemble averaged quantities was shown to work even for extremely short trajectories, where single particle analysis failed completely. The intuitive way to improve single particle characterization is to measure longer tracks. This is usually a complicated experimental challenge which demands more complex and expensive equipment. The MSD error estimation technique enables the experimentalist to improve the characterisation of the single particle distribution by measuring more particles and not longer trajectories. Thus it carries high potential to reduce experimental costs and complexity.

The deviation of the ensemble averaged NC-MSD (i.e. without heterogeneity correction) from the heterogeneity corrected MSD, can give important insight regarding the function of the system. Since the ensemble includes particles with high anomalous exponents in addition to particles with low exponents, the ensemble covers a larger area than would be expected if all particles behaved according to the average particle. At the same time, particles with low anomalous exponents tend to stay more localized around the origin of the motion. Thus the first group acts as searchers for distant targets while the second group raises the efficiency of interaction near the origin. We see therefore that a distribution of anomalous exponents can be beneficial in various sceneries of biological signalling and interactions.

In the near future, we expect biophysics labs will produce ever growing quantities of data. New tracer families, faster acquisition rates and larger quantities of tracers per experiment will enable the development of novel biophysical theories. The correct estimation of parameters from the growing quantities of data is essential for these theories to be accurate and preform as useful predictors of biological processes.

We hope that the techniques described will enable experimentalists to better characterise SPT results. This in turn will now doubt improve our understanding of biophysical systems, and help in creating better theoretical models for their description.

Materials and Methods

Simulation of FBM

Particle tracks were simulated using the MATLAB *wfbm* function. Each path i was built for N time points with a random Hurst index, H_i , taken from a distribution with mean μ_α and variance σ_α^2 (for FBM, $2H = \alpha$). The paths were then down-sampled by a factor of 16 to give T time points. This down-sampling removes high frequency errors in the data. *Wfbm* uses the algorithm developed by Sellan and Meyer [33], and thus the simulated diffusion constant of each particle obeys $D_{\alpha,i} = \Gamma(1 - \alpha) \frac{\cos(\pi\alpha/2)}{\pi\alpha/2}$. In order to simulate paths with a minimal correlation between α_i and $D_{\alpha,i}$ each path was multiplied by $\sqrt{\Gamma(1 - \alpha)\sin(\pi\alpha/2)}$. Since $\Gamma(1 - \alpha)\Gamma(1 - \alpha) \frac{\sin(\pi\alpha/2)\cos(\pi\alpha/2)}{\pi\alpha/2} = 1$ this significantly reduces the correlation. Measurement noise was simulated by adding a random noise increment $N_{i,(t)}$ to each track at each time point. This increment was taken from a stationary normal distribution with a mean of zero and a variance of ρ^2 according to the requested noise scenario.

Path analysis

Since the purpose of this article is to present new methods for the analysis of anomalous processes, the analysis of simulated data was performed as if no prior data was available. Specifically, a separate Mathematica program was written, which analysed the “experimental” paths only with the help of another set of simulated “experimental noise” paths with the same ρ^2 variance and no FBM process.

Acknowledgments

EK would like to thank Eli Barkai for many helpful discussions and directions.

References

1. Saxton MJ, Jacobson K (1997) Single-particle tracking: applications to membrane dynamics. *Annual Review of Biophysics and Biomolecular Structure* 26: 373-399.
2. Golding I, Cox EC (2004) Rna dynamics in live escherichia coli cells. *Proceedings of the National Academy of Sciences of the United States of America* 101: 11310-11315.
3. Wong IY, Gardel ML, Reichman DR, Weeks ER, Valentine MT, et al. (2004) Anomalous diffusion probes microstructure dynamics of entangled f-actin networks. *Phys Rev Lett* 92: 178101.
4. Weigel AV, Simon B, Tamkun MM, Krapf D (2011) Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proceedings of the National Academy of Sciences* 108: 6438-6443.
5. Bronstein I, Israel Y, Kepten E, Mai S, Shav-Tal Y, et al. (2009) Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys Rev Lett* 103: 018102.
6. Kues T, Peters R, Kubitscheck U (2001) Visualization and tracking of single protein molecules in the cell nucleus. *Biophysical journal* 80: 2954-2967.
7. Bouchaud JP, Georges A (1990) Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Physics Reports* 195: 127 - 293.
8. Metzler R, Klafter J (2000) The random walk’s guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports* 339: 1 - 77.
9. Saxton MJ (2007) A biological interpretation of transient anomalous subdiffusion. i. qualitative model. *Biophysical Journal* 92: 1178 - 1191.
10. Lutz E (2001) Fractional langevin equation. *Phys Rev E* 64: 051106.
11. Wang K, Tokuyama M (1999) Nonequilibrium statistical description of anomalous diffusion. *Physica A: Statistical Mechanics and its Applications* 265: 341 - 351.
12. Kou SC, Xie XS (2004) Generalized langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. *Phys Rev Lett* 93: 180603.
13. Weber SC, Spakowitz AJ, Theriot JA (2010) Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys Rev Lett* 104: 238102.
14. Lizana L, Ambjörnsson T, Taloni A, Barkai E, Lomholt MA (2010) Foundation of fractional langevin equation: Harmonization of a many-body problem. *Phys Rev E* 81: 051118.

15. Bel G, Barkai E (2005) Weak ergodicity breaking in the continuous-time random walk. *Phys Rev Lett* 94: 240602.
16. He Y, Burov S, Metzler R, Barkai E (2008) Random time-scale invariant diffusion and transport coefficients. *Phys Rev Lett* 101: 058101.
17. Bancaud A, Huet S, Daigle N, Mozziconacci J, Beaudouin J, et al. (2009) Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *The EMBO journal* 28: 3785–3798.
18. Szymanski J, Weiss M (2009) Elucidating the origin of anomalous diffusion in crowded fluids. *Physical review letters* 103: 38102.
19. Tejedor V, Bnichou O, Voituriez R, Jungmann R, Simmel F, et al. (2010) Quantitative analysis of single particle trajectories: Mean maximal excursion method. *Biophys J* 98: 1364-1372.
20. Condamin S, Tejedor V, Voituriez R, Bénichou O, Klafter J (2008) Probing microscopic origins of confined subdiffusion by first-passage observables. In: *Proc. Natl. Acad. Sci. USA*. 105, pp. 5675-5680.
21. Jeon JH, Metzler R (2010) Analysis of short subdiffusive time series: scatter of the time-averaged mean-squared displacement. *J Phys A* 43: 252001.
22. Magdziarz M, Weron A, Burnecki K, Klafter J (2009) Fractional brownian motion versus the continuous-time random walk: A simple test for subdiffusive dynamics. *Phys Rev Lett* 103: 180602.
23. Jeon JH, Metzler R (2010) Fractional brownian motion and motion governed by the fractional langevin equation in confined geometries. *Phys Rev E* 81: 021103.
24. Kepten E, Bronshtein I, Garini Y (2011) Ergodicity convergence test suggests telomere motion obeys fractional dynamics. *Phys Rev E* 83: 041919.
25. Burnecki K, Kepten E, Janczura J, Bronshtein I, Garini Y, et al. (2012) Universal algorithm for identification of fractional brownian motion. a case of telomere subdiffusion. *Biophysical Journal* 103: 1839–1847.
26. Granek R (2011) Proteins as fractals: Role of the hydrodynamic interaction. *Physical Review E* 83: 020902.
27. Weber S, Theriot J, Spakowitz A (2010) Subdiffusive motion of a polymer composed of subdiffusive monomers. *Physical Review E* 82: 011913.
28. Qian H, Sheetz M, Elson E (1991) Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophysical journal* 60: 910.
29. Saxton M (1997) Single-particle tracking: the distribution of diffusion coefficients. *Biophysical journal* 72: 1744.
30. Michalet X (2010) Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E* 82: 041914.
31. Michalet X, Berglund A (2012) Optimal diffusion coefficient estimation in single-particle tracking. *Physical Review E* 85: 061916.
32. Martin D, Forstner M, Käs J (2002) Apparent subdiffusion inherent to single particle tracking. *Biophysical journal* 83: 2109–2117.

33. Abry P (1996) The wavelet-based synthesis for fractional brownian motion proposed by f. sellan and y. meyer: Remarks and fast implementation. Applied and computational harmonic analysis 3: 377–383.

Figure Legends

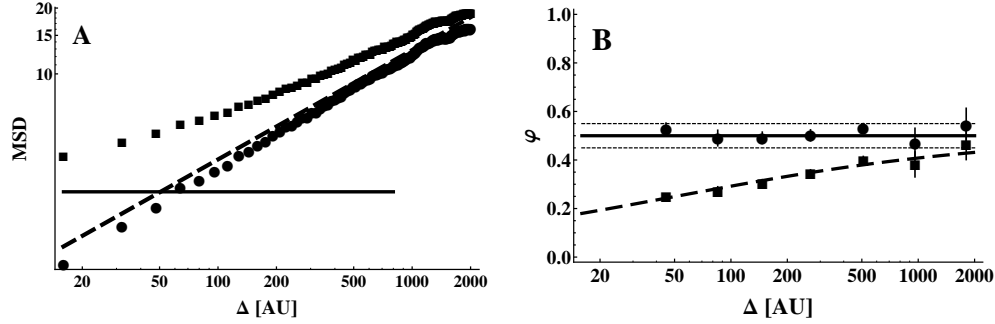


Figure 1: **Noise causes a systematic error in TAMSDs.** Trajectories of 2^{11} time points were simulated with $\alpha = 0.5$ together with an experimental measurement noise of $\rho = 1.2$. a) The TAMSD (squares) and NC-MSD (circles) of one of these trajectories are shown together with the theoretical TAMSD without measurement noise (dashed rising line). The noise TAMSD is shown as a parallel continues line. As predicted, the regular TAMSD shows an increasing error at small Δ while the NC-MSD presents a good fit. b) The dynamic functional of the TAMSD (squares) was calculated at various time points and was found to follow the predicted erroneous values. The dynamic functional of the NC-MSD (circles) on the other hand closely follows the simulated anomalous exponent (continuous line, dotted lines are $\pm 10\%$ error of α). Since a finite amount of time points is used for each TAMSD, values of φ vary slightly between individual MSDs. The values presented are averages and perpendicular lines show standard errors.

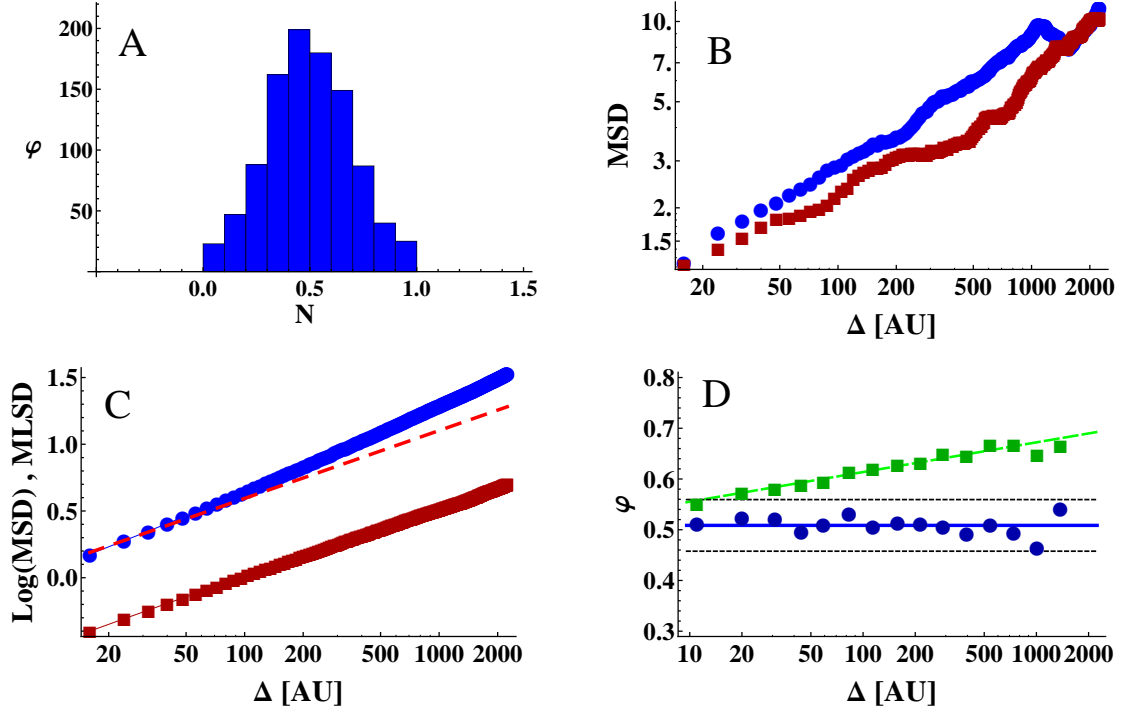


Figure 2: **Correction of the ensemble variation error.** a) For 2^{10} particles, an anomalous exponent was taken from a normal distribution with $\mu_\alpha = 0.5$ and $\sigma_\alpha = 0.2$. Total time was $T = 2^{13}$ and 2^{10} time points were taken. b) The TAMSDs of two representative particles are shown. The roughness of the path and random deviation from a straight line can be seen. c) The regular EA-TAMSD (circles, blue), the MLSD (squares, red) and the theoretical prediction for the average particle (dashed red line) can be seen. Lines in MSD and MLSD are to guide the eye. The deviating slope of the MSD can be seen while the MLSD is parallel to the theoretical average particle. d) The dynamic functional was calculated for the EA-TAMSD (squares, green) and MLSD (circles, blue) at constant intervals on the logarithmic scale. The EA-TAMSD dynamic functional is linear and rising (dashed green line), while the MLSD is around the simulated $\mu_\alpha = 0.5$ (blue line). Dotted black lines denote 10% errors from μ_α .

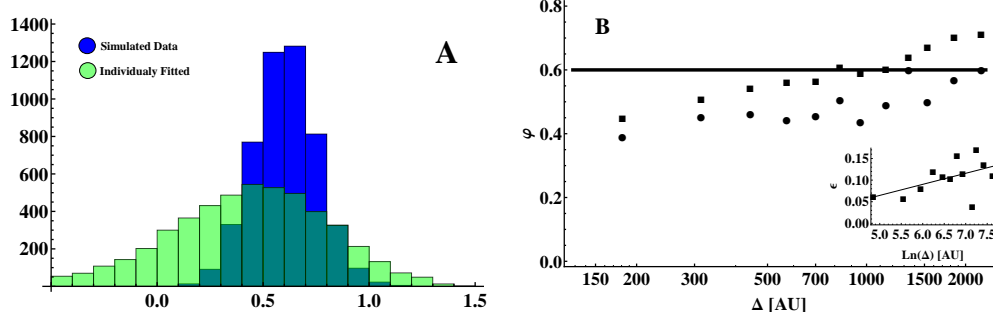


Figure 3: **Finding the standard deviation of anomalous exponents.** 5000 short trajectories (2^6 time points, down sampled from paths of length 2^{13}), where simulated with $\mu_\alpha = 0.6$ and $\sigma_\alpha = 0.15$. In addition an experimental noise of $\rho = 1.5$ was added. a) Trying to find the anomalous exponent for each track individually fails, as a much wider distribution is found (green) with an offset towards lower values, compared to the simulated values (blue). b) For 12 time points, the dynamic functional of the EA-TAMSD (square) and MLSD (circle) is calculated. Even though they are both under the simulated average (straight line), the differences can be fitted to a straight line on a logarithmic scale (inset). The resulting standard deviation found was $\sigma_\alpha = 0.16$, an error of 7%.

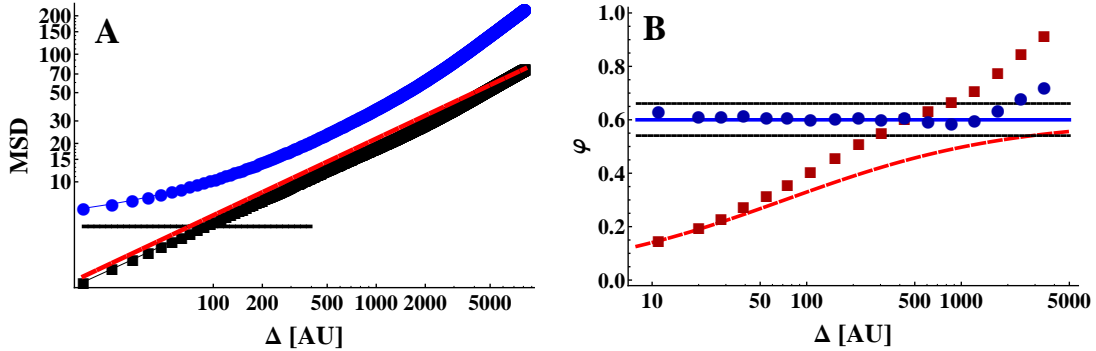


Figure 4: **Constructing the average particle MSD.** 1000 trajectories of 2^{10} time points where simulated from a distribution of anomalous exponents with $\mu_\alpha = 0.6$ and $\sigma_\alpha = 0.2$. A measurement noise of $\rho = 1.5$ was added. a) The EA-TAMSD (blue circles) shows strong deviations both in short and long times. The corrected average MSD (ν_c , black squares) is almost identical to the theoretical average particle (dashed red). b) Calculating the dynamic functional for the EA-TAMSD (red squares) and ν_c (blue circles) shows the dramatic improvement. Blue line is the ensemble average, $\mu_\alpha = 0.6$, dashed red line is the noise induced error and black dotted lines are 10% errors from μ_α